## 3.4 大数定理和中心极限定理

在数学中大家都注意到过这样的现象:有的时候一个有限的和很难求,但一经取极限由有限过渡到无限,则问题反而好办.例如,若要对某一有限范围的 $x$ 计算和

$$a_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

则在 $n$ 固定但很大时,很难求.而一经取极限,则有简单的结果:

$\lim\limits_{n\to\infty}a_n(x)=e^x$. 利用这个结果，当 $n$ 很大时，可以把 $e^x$ 作为$a_n(x)$的近似值.

在概率论中也存在着这种情况. 如果 $X_1, X_2, \cdots, X_n$ 是一些随机变量，则 $X_1+\cdots+X_n$ 的分布，除了若干例外，算起来很复杂. 因而自然地会提出问题：可否利用极限的方法来进行近似计算？事实证明这不仅可能，且更有利的是：在很一般的情况下，和的极限分布就是正态分布. 这一事实增加了正态分布的重要性. 在概率论上，习惯于把和的分布收敛于正态分布的那一类定理都叫做"中心极限定理". 在本节 3.4.2 段中我们将列述这类定理中最简单，然而也是最重要的一种情况.

在概率论中，另一类重要的极限定理是所谓"大数定理". 它是由概率的统计定义"频率收敛于概率"引伸而来. 为描述这一点，我们把频率通过一些随机变量的和表示出来. 设做了 $n$ 次独立试验，每次观察某事件 $A$ 是否发生. 按(1.20)式定义随机变量 $X_i, i=1,\cdots,n$. 则在这 $n$ 次试验中事件$A$ 一共出现了 $X_1+\cdots+X_n$ 次，而频率为

$$p_n = (X_1+\cdots+X_n)/n = \bar{X}_n \tag{4.1}$$

若 $P(A)=p$，则"频率趋于概率"就是说，在某种意义下(详见下文)，当 $n$ 很大时$p_n$ 接近$p$. 但 $p$ 就是$X_i$ 期望值，故也可以写成：

当 $n$ 很大时 $\bar{X}_n$ 接近于$X_i$ 的期望值.

按这个表述，问题就可以不必局限于 $X_i$ 只取0,1 两个值的情形. 事实也是如此. 这就是较一般情况下的大数定理. "大数"的意思，就是指涉及大量数目的观察值 $X_i$，它表明这种定理中指出的现象，只有在大量次数的试验和观察之下才能成立. 例如，一所大学可能包含上万名学生，每人有其身高. 如果我们随意观察一个学生的身高 $X_1$，则 $X_1$ 与全校学生的平均身高 $a$ 可能相去甚远. 如果我们观察 10 个学生的身高而取其平均，则它有更大的机会与 $a$ 更接近些. 如观察 100 个，则其平均又能更与 $a$ 接近些. 这些都是我们日常经验中所体验到的事实. 大数定理对这一点从理论的高

度给予概括和论证.

### 3.4.1 大数定理

**定理 4.1** 设 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量，记它们的公共均值为 $a$. 又设它们的方差存在并记为 $\sigma^2$. 则对任意给定的 $\varepsilon>0$ 有

$$\lim_{n\to\infty} P(|\bar{X}_n - a| \geqslant \varepsilon) = 0 \quad (\bar{X}_n \text{ 见(4.1)}) \tag{4.2}$$

(4.2)这个式子指出了"当 $n$ 很大时，$\bar{X}_n$ 接近 $a$"的确切含义:它的意义是概率上的，不同于微积分意义下某一列数 $a_n$ 收敛于数 $a$. 按(4.2)只是说:不论你给定怎样小的 $\varepsilon>0$, $\bar{X}_n$ 与 $a$ 的偏离有否可能达到 $\varepsilon$ 或更大呢? 这是可能的，但当 $n$ 很大时，出现这种较大偏差的可能性很小，以致当 $n$ 很大时，我们有很大的(然而不是百分之百的)把握断言 $\bar{X}_n$ 很接近 $a$. 拿上面学生身高的那个例子说，即使你抽了 100 个以至 1000 个学生，你有没有绝对的把握说，这 100 个或 1000 个学生的平均身高一定很接近全校学生的平均身高 $a$ 呢? 没有，因为理论上不能排除这种可能性:你碰巧把全校中那 100 或 1000 个最高的学生都抽出来了. 这时你计算的 $\bar{X}_n$ 就会与 $a$ 有很大差距. 但我们也能相信，如果抽样真是随机的(每一学生有同等被抽出的机会)，则随着抽样次数增多，这样的可能性会愈来愈小. 这就是(4.2)式的意思. 像(4.2)式这样的收敛性，在概率论中叫做"$\bar{X}_n$ 依概率收敛于 $a$".

为了证明定理 4.1，需要下面的概率不等式:

**马尔科夫不等式** 若 $Y$ 为只取非负值的随机变量，则对任给常数 $\varepsilon>0$ 有

$$P(Y \geqslant \varepsilon) \leqslant E(Y)/\varepsilon \tag{4.3}$$

设 $Y$ 为连续型变量，密度函数为 $f(y)$. 因为 $Y$ 只取非负值，有 $f(y)=0$ 当 $y<0$. 故

$$E(Y) = \int_0^\infty yf(y)\mathrm{d}y \geqslant \int_\varepsilon^\infty yf(y)\mathrm{d}y$$

因为在$[\varepsilon,\infty)$内总有 $y\geqslant\varepsilon$,且$\int_{\varepsilon}^{\infty}f(y)\mathrm{d}y$ 就是$P(Y\geqslant\varepsilon)$. 故

$$E(Y)\geqslant\int_{\varepsilon}^{\infty}yf(y)\mathrm{d}y\geqslant\varepsilon\int_{\varepsilon}^{\infty}f(y)\mathrm{d}y=\varepsilon P(Y\geqslant\varepsilon)$$

即(4.3). 当 $Y$ 为离散型时证明相似,请读者自己完成.

不等式(4.3)的一个重要特例为

**契比雪夫不等式**. 若 $\mathrm{Var}(Y)$存在,则

$$P(|Y-EY|\geqslant\varepsilon)\leqslant\mathrm{Var}(Y)/\varepsilon^2 \tag{4.4}$$

为证此,只须在(4.3)式中以$[Y-EY]^2$ 代 $Y$,$\varepsilon^2$ 代 $\varepsilon$,并注意$P((Y-EY)^2\geqslant\varepsilon^2)=P(|Y-EY|\geqslant\varepsilon)$即可.

现在转到定理 4.1 的证明. 利用契比雪夫不等式(4.4),并注意 $E(\bar{X}_n)=\sum_{i=1}^{n}E(X_i)/n=na/n=a$,得

$$P(|\bar{X}_n-a|\geqslant\varepsilon)\leqslant\mathrm{Var}(\bar{X}_n)/\varepsilon^2 \tag{4.5}$$

因为$\bar{X}_n=\frac{1}{n}(X_1+\cdots+X_n)$而 $X_1,\cdots,X_n$ 独立,有

$$\mathrm{Var}(\bar{X}_n)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)=\frac{1}{n^2}n\sigma^2=\sigma^2/n$$

以此代入(4.5),得

$$P(|\bar{X}_n-a|\geqslant\varepsilon)\leqslant\sigma^2/(n\varepsilon^2)\to 0,\text{当 } n\to\infty$$

这证明了(4.2).

定理 4.1 的一个重要特例,即前面提到的"频率收敛于概率":

$$\lim_{n\to\infty}P(|p_n-p|\geqslant\varepsilon)=0 \tag{4.6}$$

这个定理是最早的一个大数定理,是伯努利在 1713 年一本著作中证明的,常称为伯努利大数定理.

大数定理的研究是概率论中一个很重要、古老且至今仍尚活跃的课题,有许多深刻的结果. 例如,不用假定 $X_i$ 的方差存在也可以证明(4.2)式:$X_1,X_2,\cdots$不必同分布甚至也可以不独立(当然仍得有一定限制),收敛也可以改成其他更强的形式等. 这些都超出本书的范围之外.

在概率论中，大数定理常称为"大数定律". 这个字面上的不同，也不见得有很特殊的含义. 但是，"定理"一词往往用于指那种能用数学工具严格证明的东西，而"定律"则不一定是这样. 如牛顿的力学三大定律，电学中的欧姆定律之类. 这牵涉到一个从哪个角度去看的问题. 像(4.2)式这样有确切数学表述，并能在一定的理论框架内证明的结果，称之为"定理"无疑是恰当的. 可是，当我们泛泛地谈论"平均值的稳定性"(即稳定到理论上的期望值)时，这表述了一种全人类多年的集体经验，有些哲理的味道. 且这种意识也远早于现代概率论给之以严格表述之前，因此，称之为"定律"也不算不恰当.

### 3.4.2 中心极限定理

中心极限定理的意义已在本节开始处阐述过了. 如我们所曾指出的，这是指一类定理. 下面的定理 4.2 是其中之一：

**定理 4.2** 设 $X_1, X_2, \cdots, X_n, \cdots$ 为独立同分布的随机变量，$E(X_i) = a$, $\mathrm{Var}(X_i) = \sigma^2$, $0 < \sigma^2 < \infty$. 则对任何实数 $x$，有

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - na) \leqslant x\right) = \Phi(x) \quad (4.7)$$

这里 $\Phi(x)$ 是标准正态分布 $N(0,1)$ 的分布函数，即

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2/2} dt \quad (4.8)$$

注意 $X_1 + \cdots + X_n$ 有均值 $na$，方差 $n\sigma^2$. 故

$$(X_1 + \cdots + X_n - na)/(\sqrt{n}\sigma).$$

就是 $X_1 + \cdots + X_n$ 的标准化，即使其均值变为 0 方差变为 1，以与 $N(0,1)$ 的均值方差符合.

(4.7)告诉我们，虽则在一般情况下我们很难求出 $X_1 + \cdots + X_n$ 的分布的确切形式，但当 $n$ 很大时，可以通过 $\Phi(x)$ 给出其近似值. 例如，若已知 $a = 1$, $\sigma^2 = 4$, $n = 100$. 要求 $P(X_1 + \cdots + X_{100} \leqslant 125)$. 因 $na = 100$, $\sqrt{n}\sigma = 20$，把事件 $X_1 + \cdots + X_{100} \leqslant 125$ 改写

为$(X_1+\cdots+X_{100}-100)/20\leqslant 1.25$,用(4.7)得到上述概率的近似值为 $\Phi(1.25)=0.8944$. 这里当然有一定的误差. 有许多研究工作就是为了估计这种误差,也得出了一些深刻的结果. 但是,这种误差估计要求对 $X_i$ 的分布或其矩有一定的了解.

定理 4.2 通称为林德伯格定理或林德伯格-莱维定理,是这两位学者在本世纪 20 年代证明的."中心极限定理"的命名也是始于这个时期,它是波伊亚在 1920 年给出的. 但定理 4.2 并非最早的中心极限定理. 历史上最早的中心极限定理是定理 4.2 的一个特例,即当 $X_i$ 由(1.20)式定义时,这时,如以前多次指出的,$X_1+\cdots+X_n$ 就是某事件 $A$ 在 $n$ 次独立试验中发生的次数. 这个特例很重要,值得单独列为一条定理.

**定理 4.3** 设 $X_1,X_2,\cdots,X_n,\cdots$独立同分布,$X_i$ 分布是

$$P(X_i=1)=p,P(X_i=0)=1-p,0<p<1$$

则对任何实数 $x$,有

$$\lim_{n\to\infty}P\left(\frac{1}{\sqrt{np(1-p)}}(X_1+\cdots+X_n-np)\leqslant x\right)=\Phi(x) \tag{4.9}$$

定理 4.3 是定理 4.2 的特例,只须注意 $E(X_i)=p$,$\mathrm{Var}(X_i)=p(1-p)$. 又此处 $X_1+\cdots+X_n$ 服从二项分布 $B(n,p)$,故定理 4.3 是用正态分布去逼近二项分布. 在第二章例 1.2 曾指出过用波哇松分布逼近二项分布. 二者的应用不同:(4.9)用于 $p$ 固定,因而当 $n$ 很大时 $np$ 很大. 而波哇松逼近则用于 $p$ 很小(可设想成 $p$ 随 $n$ 变化以趋向于 0)但 $np=\lambda$ 不太大时. 共同之点是 $n$ 必须相当大.

定理 4.3 称为棣莫弗-拉普拉斯定理,是历史上最早的中心极限定理. 1716 年棣莫弗讨论了 $p=\frac{1}{2}$ 的情形,而拉普拉斯则把它推广到一般 $p$ 的情形.

如果 $t_1,t_2$ 是两个正整数,$t_1<t_2$. 则当 $n$ 相当大时,按(4.9)近似地有

$$P(t_1 \leqslant X_1 + \cdots + X_n \leqslant t_2) \approx \Phi(y_2) - \Phi(y_1) \quad (4.10)$$

其中

$$y_i = (t_i - np)/\sqrt{np(1-p)}, \quad i = 1,2 \quad (4.11)$$

我们指出:若把 $y_1, y_2$ 修正为

$$y_1 = \left(t_1 - \frac{1}{2} - np\right)/\sqrt{np(1-p)}$$

$$y_2 = \left(t_2 + \frac{1}{2} - np\right)/\sqrt{np(1-p)} \quad (4.12)$$

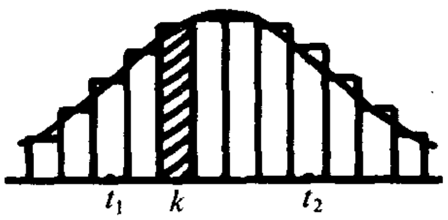

图 3.7

再应用公式(4.10),则一般可提高精度.其道理可以从图 3.7 看出.此图中每一矩形小条底边长为 1,底边中点为非负整数 $k$,而矩形的高,就是 $P(X_1 + \cdots + X_n = k)$,即二项概率 $b(k; n, p)$.图中的曲线则是正态分布 $N(np, np(1-p))$ 的密度函数的曲线.近似式(4.10)的意思,无非是用这曲线下的面积来近似代替这些矩形条的面积.可是细看图形 3.7,可知,包括点 $t_1, t_1+1, \cdots, t_2$,这些小条在横轴上所占范围,是左起 $t_1 - 1/2$,右止 $t_2 + 1/2$,故曲线下的面积,也应在这两个起止点之间去计算.这就是修正公式(4.12)的来由.当 $n$ 很大时,这个修正并不很重要,但在 $n$ 不太大时则有比较大的影响.

**例 4.1** 设某地区内原有一家小型电影院,因不敷需要,拟筹建一所较大型的.设据分析,该地区每日平均看电影者约有 $n = 1600$ 人,且预计新电影院建成开业后,平均约有 3/4 的观众将去这新影院.

现该影院在计划其座位数时,要求座位数尽可能多,但"空座达到 200 或更多"的概率又不能超过 0.1.问设多少座位为好?

设把每日看电影的人排号为 1,2⋯,1600,且令

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 个观众去新影院} \\ 0, & \text{若不然} \end{cases} \quad i = 1, \cdots, 1600$$

则按假定有 $P(X_i=1)=3/4, P(X_i=0)=1/4$. 又假定各观众去不去电影院系独立选择，则 $X_1, X_2, \cdots$ 是独立随机变量.

现设座位数为 $m$，则按要求

$$P(X_1+\cdots+X_{1600}\leqslant m-200)\leqslant 0.1$$

在这个条件下取 $m$ 最大. 这显然就是在上式取等号时，因为 $np=1600\cdot(3/4)=1200, \sqrt{np(1-p)}=10\sqrt{3}$，按(4.12)的修正，$m$ 应满足条件

$$\Phi\left(\left(m-200+\frac{1}{2}-1200\right)\Big/(10\sqrt{3})\right)=0.1$$

查 $\Phi(x)$ 的表得知，当 $\Phi(x)=0.1$ 时，$x=-1.2816^*$. 由

$$(m-200+1/2-1200)\big/(10\sqrt{3})=-1.2816$$

定出 $m=1377.31\approx 1377$. 在本例中，(4.12)式的修正没有什么影响.

直到本世纪 30 年代，中心极限定理的研究曾是概率论的一个重要内容，至今仍是一个活跃的方向. 推广的方向如独立不同分布以至非独立的情形，由中心极限定理而引起的误差的估计，以及与之相关联的问题如大偏差问题之类.